USPTO Serial Number: 10/014,744
Nong Ye *et al.*
Reply to Office Action dated February 9, 2004

Amendment to the Specification:

Replace paragraph [0001] with the following amended paragraph:

Field of the Invention

[0001] The present invention relates in general to a method of classifying data points, and more particularly, to a method of incrementally learning classification patterns and using those patterns to classify data points.

C1

Replace paragraph [0002] with the following amended paragraph:

Background of the Invention

[0002] The general concept of classifying data points has been used in a myriad of contexts and applications. In a signature recognition application, a group of data points must be classified in order to identify a particular pattern. A signature recognition system using data classification techniques can identify a particular human face from a crowd, regulate the flow of inventory in a manufacturing system, or perform medical diagnosis from patient data. In computer technology, classification of data points can be used for intrusion detection and computer security. An intrusion can be defined as any set of activities aimed at breaking the security of a computer network system. Intrusions may take many forms: external attacks, internal misuses, network-based attacks,

C2

information gathering, denial of service, etc. Information security against intrusions mainly involves intrusion prevention, detection, diagnosis, response, and system recovery stages.

C2

Replace paragraph [00046] with the following amended paragraph:

[0003] In step 24 of FIG. 3, distance measures are determined between the first data and a plurality of clusters which include the dummy cluster. For the data points X defined in Table 1, the nearest cluster L to each data point is determined using the distance measure, d(X,L), weighted by the squared correlation coefficient of each predictor variable. A cluster L is a summarization of the data points in it, and is represented by the centroid of all the data points in it, with coordinates XL, the number of data points, N_L, and the class label, XL_T. The distance measure can be calculated by equation (11) using a weighted Canberra distance, or by equation (12) using a weighted Euclidean distance, or by equation (13) using a weighted Chi-squared distance.

C3

$$d(X,L) = \sum_{i=1}^{P} \frac{|X_i - XL_i|}{X_i + XL_i} r_{iT}^2 \qquad (11)$$

$$d(X,L) = \sqrt{\sum_{i=1}^{P} (X_i - XL_i)^2 r_{iT}^2} \qquad (12)$$

C3

$$d(X,L)=\sum_{i=1}^{p}\frac{(X_i-XL_i)^2}{XL_1}r_{iT}^2 \qquad (13)$$